

Nonpositive curvature, the variance functional, and the Wasserstein barycenter *

Young-Heon Kim[†] and Brendan Pass[‡]

March 24, 2015

Abstract

This paper connects nonpositive sectional curvature of a Riemannian manifold with the displacement convexity of the variance functional on the space $P(M)$ of probability measures over M . We show that M has nonpositive sectional curvature and has trivial topology (i.e. is homeomorphic to \mathbb{R}^n) if and only if the variance functional on $P(M)$ is displacement convex. This is followed by a Jensen type inequality for the variance functional with respect to Wasserstein barycenters, as well as by a result comparing the variance of the Wasserstein and linear barycenters of a probability measure on $P(M)$ (that is, an element of $P(P(M))$). These results are applied to invariant measures under isometry group actions, giving a comparison for the variance functional between the Wasserstein projection and the L^2 projection to the set of invariant measures.

1 Introduction

In this paper, we study the influence of nonpositive sectional curvature of a complete Riemannian manifold M on the geometry of the space $P(M)$ of probability measures, equipped with the Wasserstein metric.

Given a probability measure μ on M , the variance of μ is defined by

$$\text{var}(\mu) := \inf_{y \in M} \int_M d^2(x, y) d\mu(x),$$

where d denotes the Riemannian distance. A minimizer $y \in M$ in the above is often called a barycenter of μ . We are interested in the way that the variance,

*Y.-H.K. is supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants 371642-09 and 2014-0544, as well as Alfred P. Sloan research fellowship 2012-2016. B.P. is pleased to acknowledge the support of a University of Alberta start-up grant and National Sciences and Engineering Research Council of Canada Discovery Grant number 412779-2012.

[†]Department of Mathematics, University of British Columbia, Vancouver BC Canada V6T 1Z2 yhkim@math.ubc.ca

[‡]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1 pass@ualberta.ca.

viewed as a functional on the space $P(M)$ of Borel probability measures on M , interacts with the geometry on $P(M)$ induced by the Wasserstein distance; the Wasserstein distance between $\mu, \nu \in M$ is given by

$$W_2(\mu, \nu) := \inf_{\pi_{\#}^1 \gamma = \mu, \pi_{\#}^2 \gamma = \nu} \int_{M \times M} d^2(x, y) d\gamma(x, y), \quad (1.1)$$

where, for $i = 1, 2$, $\pi_{\#}^i \gamma$ denotes the pushforward of γ by the canonical projections, $\pi^1(x, y) = x$, $\pi^2(x, y) = y$, respectively. Recall that in general, the pushforward $T_{\#} \sigma$ of a measure σ by a map $T : X \rightarrow Y$, is defined by $T_{\#} \sigma(A) := \sigma(T^{-1}(A))$ for all measurable sets $A \subset Y$.

We will show that the combination of nonpositive sectional curvature together with trivial topology, is characterized by displacement convexity of the variance; that is, convexity along geodesics on $P(M)$ induced by the Wasserstein metric (see Theorem 2.1 below). The notion of displacement interpolation, initiated by McCann [11], gives a natural geometric way to interpolate between two probability measures. In turn, convexity of certain functionals with respect to this interpolation, known as displacement convexity, has proven to be a remarkably powerful tool in proving geometric and functional inequalities, and has found applications in physics and economics as well; see, e.g. [19, 20].

Let us note that there are already many known characterizations of non-positive sectional curvature; in fact there is one involving the variance functional, due to Sturm [18, Theorem 4.9], which applies to more general spaces than we consider here. We believe, however, that it is interesting to have a characterization involving displacement convexity, particularly in light of the now well known characterization of Ricci curvature bounds involving displacement convexity of the entropy functional, developed by many authors, including Cordero-Erausquin-McCann-Schmuckenschlager [4], Otto-Villani [13] and Sturm-Von-Renesse [21], and culminating in the recent work of Lott-Villani [9] and Sturm [16, 17].

Note that, unlike many other interesting displacement convex functionals, the variance functional is well defined and finite as soon as the measure has finite second moment (ie, one does not require absolute continuity with respect to volume), and is weak-* continuous. This property makes it particularly well suited for studying sectional curvature bounds. Heuristically, displacement interpolation moves a measures along a family of non-intersecting geodesics with fixed endpoints. Nonnegative Ricci curvature tends to pull those geodesics apart at intermediate times; this is quantified by the displacement convexity of the entropy in the works cited above. Our setting is slightly different; we expect nonpositive sectional curvature to contract geodesics at intermediate times in a certain sense. However, as sectional curvature is a property of two dimensional sections of the tangent space, this contraction may not be detectable by functionals which are finite only on absolutely continuous measures. For instance, if the sectional curvature of some section is positive, but the Ricci curvature is everywhere negative, the volume of a small ball will get contracted. However, a set which is concentrated and interpolated along the directions with

positive sectional curvature can get spread out in a certain sense; the variance turns out to be an appropriate way to quantify this.

We go on to extend the convexity of the variance to convexity with respect to Wasserstein barycenters: see Theorem 3.6. Analogously to the definition of barycenters of measures on M , a barycenter $BC^W(\Omega)$ of a measure Ω on $P(M)$, which we call a Wasserstein barycenter, or simply, W_2 -barycenter, of Ω is defined as a minimizer of

$$\nu \mapsto \int_{P(M)} W_2^2(\mu, \nu) d\Omega(\mu). \quad (1.2)$$

The notion of Wasserstein barycenters was considered by Agueh-Carlier [1] when M is a subset in the Euclidean space $M \subseteq \mathbb{R}^n$ and Ω is a discrete measure on $P(M)$, and later by the present authors [8] for Riemannian manifolds M and general probability measures Ω on $P(M)$. It extends displacement interpolation, allowing one to interpolate between several (or, in our formulation, even infinitely many) probability measures in a canonical way. Agueh and Carlier [1] also considered convexity over Wasserstein barycenters, as a generalization of displacement convexity. This notion can be interpreted as an analogue of Jensen's inequality; this point of view was investigated in [8] where geometric versions of Jensen's inequality were established for displacement convex functionals on Wasserstein spaces over Riemannian manifolds, extending the Euclidean results of [1].

The displacement convexity of the variance should be contrasted with its behaviour with respect to linear interpolation of measures. When measures are interpolated linearly, it is easy to see that the variance is concave, regardless of the curvature of M . Combined with the ordinary Jensen's inequality and our displacement convexity result, this implies that the variance of the Wasserstein barycenter of any measure Ω on $P(M)$ is less than or equal to the variance of its linear barycenter, if M is nonpositively curved simply connected space; see Corollary 4.1. Although this statement is not explicitly linked to convexity and concavity, we are not aware of another proof which does not use convexity over Wasserstein barycenters. We present a counterexample demonstrating that this inequality can fail when the curvature conditions are relaxed.

We then turn our attention to the special case when the measure Ω is induced by a left invariant measure on an isometry group G acting on M , and relate our work to the W_2 projection $P_G^W(\mu)$ of $\mu \in P(M)$ to the set of G -invariant measures on M . Connections between optimal transport problems and measures which are invariant under certain operations have recently begun to attract considerable attention; see [12] [22] [7] [5] [6], although these works are primarily concerned with finding Kantorovich solutions of the optimal transport problem with certain symmetry constraints, rather than looking at Wasserstein projections. Our work here implies a comparison result for the variance functional between the L^2 projection and the W_2 projection to the G -invariant set. Namely, when M is nonpositively curved and simply connected, we get, under

suitable conditions on μ ,

$$\text{var}(P_G^W(\mu)) \leq \text{var}(\mu);$$

see Corollary 5.3. Note that, at first glance, this inequality has no obvious connection to the barycenter of a family of measures, but we are not aware of another simple proof of it. Furthermore, it is interesting when contrasted with the inequality

$$\text{var}(P_G^{L^2}(\mu)) \geq \text{var}(\mu),$$

for the L^2 projection $P_G^{L^2}(\mu)$ of μ onto the G -invariant set; see (5.1).

The paper is organized as follows: In Section 2 we establish the equivalence, on complete Riemannian manifolds, between nonpositive sectional curvature, together with simple connectedness, and displacement convexity of the variance. In Section 3, we show that this displacement convexity extends to convexity over Wasserstein barycenters. Section 4 is devoted to the comparison of the behaviour of the variance functional between linear and Wasserstein barycenters. Finally, in Section 5, these results are applied to isometry group actions, yielding comparison results for the L^2 the W_2 projections to the set of invariant measures.

2 Displacement convexity of the variance and nonpositive sectional curvature

Before stating the main theorem of this section, we develop some notation. A well known result of Brenier [3] and McCann [10] asserts that if the measure μ is absolutely continuous with respect to volume and both μ and ν have finite variance, then there exists a unique minimizer γ to the minimization problem (1.1), and furthermore, $\gamma = (Id, F)_\# \mu$, where $F : M \rightarrow M$, is the unique mapping such that $F_\# \mu = \nu$ taking the form $F(x) = \exp_x(-\nabla \phi(x))$, where $\phi : M \rightarrow \mathbb{R}$ is a $\frac{d^2}{2}$ -convex function; that is, ϕ takes the form

$$\phi(x) = \sup_{y \in M} -\frac{d^2(x, y)}{2} - \phi^c(y)$$

for some $\phi^c : M \rightarrow \mathbb{R}$. The displacement interpolant between μ and ν is then the map $[0, 1] \rightarrow P(M)$ given by $\mu_t = ((1-t)\text{Id} + tDu(x))_\# \mu$. We note that this notion of displacement interpolation can be extended to non-absolutely continuous measures in $P(M)$; for precise definitions, we refer the reader to the books [2, 20]. A functional $\mathcal{F} : P(M) \rightarrow \mathbb{R} \cup \infty$ is called displacement convex if the function $t \mapsto \mathcal{F}(\mu_t)$ is convex for every displacement interpolant μ_t .

This section is then devoted to the proof of the following result:

Theorem 2.1. *Assume M is simply connected. Then M has nonpositive sectional curvature if and only if the variance functional is displacement convex.*

Proof. This follows from Theorem 2.2 and Corollary 2.8 below. \square

2.1 Displacement convexity of variance: necessary condition

In this subsection, a standard argument shows that if the variance is displacement convex, then the underlying Riemannian manifold has to be simply connected and nonpositively curved.

Theorem 2.2. *Let M be a complete Riemannian manifold. Suppose that variance is displacement convex, i.e. $\text{var}(\mu_t) \leq (1-t)\text{var}(\mu_0) + t\text{var}(\mu_1)$ for each displacement interpolation μ_t of probability measures on M . Then, M is simply connected and has nonpositive sectional curvature $K \leq 0$.*

Proof. We first tackle the simple connectedness. The proof is by contradiction; assume M is not simply connected. We claim that this implies that each point x has a nonempty cut locus. To see this, note that there are homotopically nontrivial loops from x to itself. Taking an arc-length minimizing sequence of such loops, and noting that each loop in the sequence remains in a compact subset of M , we can pass to a convergent subsequence and obtain a geodesic loop from x to itself. A cut locus point clearly exists along such a loop.

By [4, Proposition 2.5], then, for any $x \in M$, there exists $y \in M$, and a small $v \in T_x M$ such that

$$d^2(\exp_x v, y) + d^2(\exp_x(-v), y) - 2d^2(x, y) < 0. \quad (2.1)$$

Now, take two measures $\mu_0 = \frac{1}{2}[\delta_y + \delta_{\exp_x v}]$ and $\mu_1 = \frac{1}{2}[\delta_y + \delta_{\exp_x(-v)}]$. The displacement interpolant at $t = \frac{1}{2}$ is clearly $\mu_{1/2} = \frac{1}{2}[\delta_y + \delta_x]$. Note that the variances of the doubly supported measures μ_0 , μ_1 and $\mu_{1/2}$ are, respectively, $\frac{1}{2}d^2(\exp_x v, y) + \frac{1}{2}d^2(\exp_x(-v), y)$ and $\frac{1}{2}d^2(x, y)$. This contradicts the displacement convexity of the variance. We note that one could also use this to construct an example with absolutely continuous μ_0 and μ_1 ; observe that weak-* density of absolutely continuous probability measures, the weak-* continuity of the variance functional and stability of the displacement interpolation (these latter two facts are straightforward to prove; see Lemmas (3.3) and (3.2) in the next section), combined with inequality (2.1), we can find absolutely continuous measures μ_0 and μ_1 , whose displacement interpolant $\mu_{1/2}$ satisfies

$$\text{var}(\mu_1) + \text{var}(\mu_0) < 2\text{var}(\mu_{1/2}).$$

This again violates the displacement convexity of the variance, yielding the desired contradiction and therefore establishing the simple connectedness of M .

We now turn to the sectional curvature assertion. The proof is again by contradiction; assume a section Σ of a tangent space $T_x M$ has positive sectional curvature. Then, we can find, for some small $\epsilon > 0$, points x_0, x_1, y_0, y_1 with the following properties:

$$\begin{aligned} d(x_0, x_1) &= d(y_0, y_1) := \epsilon \\ d(\gamma_0(t), \gamma_1(t)) &> \epsilon \text{ for some } t \in (0, 1) \\ d^2(x_0, y_0) + d^2(x_1, y_1) &\leq d^2(x_0, y_1) + d^2(x_1, y_0) \end{aligned}$$

Here $\gamma_0(t)$ and $\gamma_1(t)$ are geodesics from x_0 to y_0 and x_1 to y_1 , respectively. Now, consider optimal transport between the two measures $\mu_0 = \frac{1}{2}[\delta_{x_0} + \delta_{x_1}]$ and $\mu_1 = \frac{1}{2}[\delta_{y_0} + \delta_{y_1}]$; the optimal plan clearly pairs x_0 with y_0 and x_1 with y_1 , and so the displacement interpolant at t is $\mu_t = \frac{1}{2}[\delta_{\gamma_0(t)} + \delta_{\gamma_1(t)}]$. Therefore, the variances of μ_0, μ_1 and $\mu_{1/2}$ are, respectively, $\frac{1}{4}d^2(x_0, x_1)$, $\frac{1}{4}d^2(y_0, y_1)$ and $\frac{1}{4}d^2(\gamma_0(t), \gamma_1(t))$, and so

$$\begin{aligned} \text{var}(\mu_t) &= \frac{d^2(\gamma_0(t), \gamma_1(t))}{4} > \frac{\epsilon^2}{4} \\ &= (1-t)\frac{d^2(x_0, x_1)}{4} + t\frac{d^2(y_0, y_1)}{4} \\ &= (1-t)\text{var}(\mu_0) + t\text{var}(\mu_1). \end{aligned}$$

This contradicts displacement convexity; if one wants to consider absolutely continuous measures instead, then, one can argue as before using an approximation argument. This contradiction yields the desired result. \square

2.2 Displacement convexity of variance: sufficient condition

The goal of this subsection is to show that variance, as a functional on the space of probability measures, is displacement convex if the underlying domain or manifold M is simply connected and has nonpositive curvature; if our domain is not complete, then we further assume that it is geodesically convex. Here, by geodesic convexity of M , we mean that for any given two points in M , any minimizing geodesics connecting these two points remains in M .

In fact, we prove convexity along a slightly more general family of paths than displacement interpolations or equivalently, W_2 -geodesics in $P(M)$. All results in this section are obtained using standard calculations and results in Riemannian geometry.

Definition 2.3 (W_2 -quasi-geodesic). *Let V be a measurable vector field defined a.e. on a Riemannian manifold M . Define for $t \in [0, 1]$, a measurable mapping T_t as $T_t(x) = \exp_x(tV(x))$ for a.e. x . Then, for each absolutely continuous probability measure μ on M , we call the 1-parameter family $\mu_t = T_{t\#}\mu$ a W_2 -quasi-geodesic.*

Notice when V is given by $\nabla\phi$ for some c -convex function ϕ , (see [10]), W_2 quasi-geodesics become W_2 geodesics. It is convenient at this point to observe a simple technical fact, which we won't need in this section but will be used in the proof of Theorem 3.6 in the following section.

Lemma 2.4 (A first variation along W_2 quasi-geodesic). *Let μ be an absolutely continuous probability measure on M and let $\mu_t = T_{t\#}\mu$ be a W_2 quasi-geodesic, given by the measurable vector field V , $T_t(x) = \exp_x tV(x)$ for a.e. x . Let $\gamma(t)$*

be a differentiable curve in M such that $\gamma(0)$ is a barycenter of μ . Then

$$\left. \frac{d}{dt} \right|_{t=0} W_2^2(\delta_{\gamma(t)}, \mu(t)) = -2 \int_M \langle \exp_x^{-1}(\gamma(0)), V(x) \rangle d\mu(x)$$

where $\langle \cdot, \cdot \rangle$ is the Riemannian inner product. Notice that $\exp_x^{-1}(\gamma(0))$ is defined whenever x is not in the cut-locus of $\gamma(0)$, which is almost every x , and thus μ -a.e. for the absolutely continuous measure μ .

Proof. First note that

$$\nabla_x \frac{1}{2} d^2(w, x) = -\exp_x^{-1}(w) \quad (2.2)$$

for a.e. x . We then have

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} W_2^2(\delta_{\gamma(t)}, \mu_t) &= \int_M \left. \frac{d}{dt} \right|_{t=0} d^2(\gamma(t), T_t(x)) d\mu(x) \\ &= \int_M \langle \nabla_w \big|_{w=\gamma(0)} d^2(w, T_0(x)), \gamma'(0) \rangle d\mu(x) \\ &\quad + \int_M \langle \nabla_y \big|_{y=T_0(x)} d^2(\gamma(0), y), T'_0(x) \rangle d\mu(x) \\ &= \left\langle \int_M \nabla_w \big|_{w=\gamma(0)} d^2(w, T_0(x)) d\mu(x), \gamma'(0) \right\rangle \\ &\quad + \int_M \langle \nabla_y \big|_{y=T_0(x)} d^2(\gamma(0), y), T'_0(x) \rangle d\mu(x) \end{aligned}$$

Note that in the calculations both above and below, we use the absolute continuity of the measure μ so that the non differentiability points of the distance squared do not effect the calculations. Now,

$$\int_M \nabla_w \big|_{w=\gamma(0)} d^2(w, T_0(x)) d\mu(x) = \nabla_w \big|_{w=\gamma(0)} \int_M d^2(w, x) d\mu(x)$$

vanishes because $\gamma(0)$ is a barycenter of μ . The result now follows from (2.2) and the observation that $T'_0(x) = V(x)$. \square

Note that convexity along W_2 quasi-geodesics implies displacement convexity. Below we will show that the variance functional $\mu \mapsto \text{var}(\mu)$ is convex along W_2 quasi-geodesics.

We will need a simple consequence of the second variation formula of arc-length; the following Lemma is a special case of a result of Sturm [18, Corollary 2.5] and so we omit the proof.

Lemma 2.5 (Convexity of distance squared for points along two geodesics). *Let M be a simply connected manifold with nonpositive sectional curvature $K \leq 0$. Let $z, w : [0, 1] \rightarrow M$ be two geodesics. Then $t \mapsto d^2(z(t), w(t))$ is convex.*

Remark 2.6. *This results can be easily extended to the case when M is a geodesically convex domain in a complete Riemannian manifold with nonpositive curvature.*

Now, we prove the main theorem of this section:

Theorem 2.7 (Convexity along W_2 quasi-geodesics). *Let M be a geodesically convex domain in a complete simply connected manifold with nonpositive sectional curvature $K \leq 0$. Let μ_t , $a \leq t \leq b$, be a W_2 quasi-geodesic in $P(M)$. Let $t \in [a, b] \rightarrow w(t) \in M$ be a geodesic. Then, $W_2^2(\delta_{w(t)}, \mu_t)$ is convex in t .*

Proof. This is an easy corollary of Lemma 2.5 and Remark 2.6. The details follow. Note that $\mu_t = T_{t\#}\mu$ where for a.e. x , $T_t(x) = \exp_x tV(x)$ for some vector field V on M .

We now observe

$$W_2^2(\delta_{w(t)}, \mu_t) = \int_M d^2(w(t), z) d\mu_t(z) = \int_M d^2(w(t), T_t(x)) d\mu_0(x).$$

where the first equality is from the definition of W_2 distance and the second equality is from $\mu_t = T_{t\#}\mu_0$. Therefore,

$$\frac{d^2}{dt^2} W_2^2(\delta_{w(t)}, \mu_t) = \int_M \frac{d^2}{dt^2} d^2(w(t), T_t(x)) d\mu_0(x).$$

Now, note that for a fixed x , $t \in [a, b] \rightarrow T_t(x)$ is a geodesic, and so using Lemma 2.5 and Remark 2.6, we see $\frac{d^2}{dt^2} d^2(w(t), T_t(x)) \geq 0$. Thus,

$$\frac{d^2}{dt^2} W_2^2(\delta_{w(t)}, \mu_t) \geq 0.$$

This completes the proof. \square

From Theorem 2.7, convexity of the variance follows immediately. The following corollary, together with Theorem 2.2 establishes Theorem 2.1.

Corollary 2.8 (Convexity of variance along W_2 quasi-geodesics). *Adopt the notation and assumptions of Theorem 2.7. Then $\text{var}(\mu_t)$ is convex in t .*

Proof. For each interval $[\alpha, \beta] \subset [a, b]$, choose a geodesic $t \in [\alpha, \beta] \rightarrow M$ with $w(\alpha)$, $w(\beta)$ being the barycenter points of μ_α , μ_β , respectively. Apply Theorem 2.7 to get convexity of $W_2^2(\delta_{w(t)}, \mu_t)$, and note that $\text{var}(\mu_t) \leq W_2^2(\delta_{w(t)}, \mu_t)$, with equality at α and β . This establishes the convexity of $\text{var}(\mu_t)$ in t . \square

3 Convexity of the variance functional with respect to W_2 barycenters.

In this section, we use convexity along W_2 quasi-geodesics to prove a convexity result with respect to W_2 barycenters (see Theorem 3.6 below); recall that W_2

barycenters were defined in (1.2). Note that Theorem 3.6 requires no regularity (ie, absolute continuity) of the measures μ in $\text{spt}(\Omega)$. Under suitable regularity conditions on Ω (see, for example, case 1 of the proof), the argument is a straightforward variant of the proof of a similar result (for different displacement convex functionals) in [8]. Much of the work in this section is related to the extension to singular measures (in which case the barycenter itself can be singular and non unique).

Throughout this section, we will assume that M is a compact domain in a Riemannian manifold. Existence of a W_2 barycenter of a probability measure Ω on $P(M)$ is easy to show. The W_2 barycenter is not generally unique; that is, there may be multiple minimizers in (1.2). However, uniqueness is known under a mild structural condition on Ω :

Proposition 3.1. *Assume M is compact (e.g. a compact domain in a manifold) and that $\Omega(P_{ac}(M)) > 0$. Then there exists a unique W_2 barycenter.*

The proof can be found in [14] and [8]. We will also need the following stability result.

Lemma 3.2. *Assume M is compact and suppose the probability measures Ω^N on $P(M)$ converge in the weak-* topology, to Ω (with respect to the W_2 -distance on $P(M)$). Then the limit of any weakly-* convergent subsequence $\bar{\mu}^N$ of barycenters of the Ω^N is a barycenter of Ω .*

Proof. The proof is a standard argument. Choose a weakly convergent subsequence, $\bar{\mu}^N \rightarrow \bar{\mu}$. For any $\mu \in P(M)$, we have

$$\begin{aligned} W_2^2(\mu, \bar{\mu}) &\leq (W_2(\mu, \bar{\mu}^N) + W_2(\bar{\mu}^N, \bar{\mu}))^2 \\ &= W_2^2(\mu, \bar{\mu}^N) + W_2^2(\bar{\mu}^N, \bar{\mu}) + 2W_2(\mu, \bar{\mu}^N)W_2(\bar{\mu}^N, \bar{\mu}) \end{aligned} \quad (3.1)$$

Integrating against Ω^N , we have

$$\begin{aligned} \int_{P(M)} W_2^2(\mu, \bar{\mu}) d\Omega^N(\mu) &\leq \int_{P(M)} W_2^2(\mu, \bar{\mu}^N) d\Omega^N(\mu) + W_2^2(\bar{\mu}^N, \bar{\mu}) \\ &\quad + 2W_2(\bar{\mu}^N, \bar{\mu}) \int_{P(M)} W_2(\mu, \bar{\mu}^N) d\Omega^N(\mu) \end{aligned} \quad (3.2)$$

Therefore, for any $\nu \in P(M)$, we have, by definition of the barycenter $\bar{\mu}^N$,

$$\begin{aligned} \int_{P(M)} W_2^2(\mu, \bar{\mu}) d\Omega^N(\mu) &\leq \int_{P(M)} W_2^2(\mu, \nu) d\Omega^N(\mu) + W_2^2(\bar{\mu}^N, \bar{\mu}) \\ &\quad + 2W_2(\bar{\mu}^N, \bar{\mu}) \int_{P(M)} W_2(\mu, \bar{\mu}^N) d\Omega^N(\mu) \end{aligned} \quad (3.3)$$

Now, as weak-* convergence is equivalent to Wasserstein convergence, $W_2(\bar{\mu}^N, \bar{\mu})$ tends to zero as $N \rightarrow \infty$, and as the term $W_2(\mu, \bar{\mu}^N)$ is uniformly bounded by the compactness of M , the last two terms on the right hand side go to zero.

By weak convergence of the Ω^N , and continuity of $\mu \mapsto W_2^2(\mu, \bar{\mu})$ and $\mu \mapsto W_2^2(\mu, \nu)$, the above inequality tends to

$$\int_{P(M)} W_2^2(\mu, \bar{\mu}) d\Omega(\mu) \leq \int_{P(M)} W_2^2(\mu, \nu) d\Omega(\mu) \quad (3.4)$$

As ν is arbitrary, this completes the proof. \square

Another standard argument shows:

Lemma 3.3. *Assume M is compact. The mapping $\mu \mapsto \text{var}(\mu)$ is continuous on $P(M)$ with respect to the weak-* topology.*

Proof. Suppose $\mu^N \rightarrow \mu$ in the weak-* topology. It is easy (in fact, almost identical to the proof in the preceding Lemma) to show that any convergent subsequence x^N of barycenters of the μ^N converges to a barycenter x of μ . We then need to show $\int_M d^2(y, x^N) d\mu^N(y) \rightarrow \int_M d^2(y, x) d\mu(y)$. We have

$$\begin{aligned} & \left| \int_M d^2(y, x^N) d\mu^N(y) - \int_M d^2(y, x) d\mu(y) \right| \\ & \leq \left| \int_M d^2(y, x^N) d\mu^N(y) - \int_M d^2(y, x) d\mu^N(y) \right| \\ & \quad + \left| \int_M d^2(y, x) d\mu^N(y) - \int_M d^2(y, x) d\mu(y) \right| \end{aligned}$$

As $N \rightarrow \infty$, the second term in the right hand side above goes to zero by weak convergence. The first term can be written as

$$\begin{aligned} & \left| \int_M [d(y, x^N) + d(y, x)] [d(y, x^N) - d(y, x)] d\mu^N(y) \right| \\ & \leq \int_M |[d(y, x^N) + d(y, x)] [d(y, x^N) - d(y, x)]| d\mu^N(y) \\ & \leq 2\text{diam}(M) \int_M |d(y, x^N) - d(y, x)| d\mu^N(y) \\ & \leq 2\text{diam}(M) \int_M d(x, x^N) d\mu^N(y) \\ & = 2\text{diam}(M) d(x, x^N) \end{aligned}$$

The result follows. \square

Corollary 3.4. *Assume M is compact. Suppose the measures Ω^N on $P(M)$ converge weakly to Ω . Then*

$$\int_{P(M)} \text{var}(\mu) d\Omega^N(\mu) \rightarrow \int_{P(M)} \text{var}(\mu) d\Omega(\mu)$$

Proof. This is an immediate consequence of the continuity of $\mu \mapsto \text{var}(\mu)$ (Lemma 3.3) and the definition of weak convergence. \square

Before we prove the main theorem of this section, we make the following observation:

Lemma 3.5. *Suppose $\bar{\mu}$ is a barycenter of the measure Ω on $P(M)$. Then $\bar{\mu}$ is the unique barycenter of $\frac{1}{2}\delta_{\bar{\mu}} + \frac{1}{2}\Omega$.*

Proof. For any ν , we have

$$\begin{aligned} \frac{1}{2}W_2^2(\nu, \bar{\mu}) + \frac{1}{2} \int_{P(M)} W_2^2(\mu, \nu) d\Omega(\mu) &\geq \frac{1}{2}W_2^2(\bar{\mu}, \bar{\mu}) + \frac{1}{2} \int_{P(M)} W_2^2(\mu, \bar{\mu}) d\Omega(\mu) \\ &= \frac{1}{2} \int_{P(M)} W_2^2(\mu, \bar{\mu}) d\Omega(\mu), \end{aligned}$$

with equality if only if $\nu = \bar{\mu}$. \square

Now we state and prove the main theorem of this section.

Theorem 3.6 (Convexity of variance with respect to the barycenter). *Assume that M is a compact, geodesically convex domain in a complete nonpositively curved manifold: thus, $\mu \mapsto \text{var}(\mu)$ is convex along W_2 quasi-geodesics. Let Ω be a Borel probability measure on $P(M)$. Let $\bar{\mu} \in P(M)$ be a W_2 barycenter of Ω . Then, we have*

$$\text{var}(\bar{\mu}) \leq \int_{P(M)} \text{var}(\mu) d\Omega(\mu).$$

Proof. The proof is divided into three, successively more general cases.

Case 1: The measure $\Omega = \sum_{i=1}^N \lambda_i \delta_{\mu_i}$ has finite support and one of the μ_i is absolutely continuous with respect to volume.

In this case, from the result of [15], the barycenter $\bar{\mu}$ is unique and absolutely continuous with respect to the volume measure (this also holds without the curvature assumption [8]). We will need to set up some relevant notation. Let T^i be the optimal map from $\bar{\mu}$ to μ^i ; by the Brenier-McCann theorem (see [10]), for a.e. x , $T^i(x) = \exp_x \nabla \phi^i(x)$ for some $d^2/2$ convex function ϕ^i . Moreover, by a straightforward adaptation of a result of Agueh-Carlier [1] (see also [8] for more general cases), we have

$$\sum_{i=1}^N \lambda_i \nabla \phi^i(x) = 0 \quad \text{for a.e. } x \in \text{spt}(\bar{\mu}). \quad (3.5)$$

For each i , let w^i be a barycenter of μ^i ; that is, $W_2^2(\delta_{w^i}, \mu^i) = \text{var}(\mu^i)$. Let \bar{w} be a barycenter of $\bar{\mu}$.

Let $\mu_t^i = T_{t\#}^i \bar{\mu} = \exp_x t \nabla \phi^i(x) \# \bar{\mu}$ be the corresponding displacement interpolations (which are of course W_2 quasi-geodesics); note that $\mu_0^i = \bar{\mu}$, $\mu_1^i = \mu^i$. Consider the geodesic $t \in [0, 1] \mapsto \gamma^i(t) \in M$ with $\gamma^i(0) = \bar{w}$, $\gamma^i(1) = w^i$, and the function

$$t \mapsto \Phi(t) = \sum_{i=1}^N \lambda_i W_2^2(\delta_{\gamma^i(t)}, \mu_t^i).$$

Use (3.5) to compute

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} \Phi(t) &= \sum_{i=1}^N \lambda_i \left. \frac{d}{dt} \right|_{t=0} W_2^2(\delta_{\gamma^i(t)}, \mu_t^i) \\ &= \sum_{i=1}^N \lambda_i \left(-2 \int_M \langle \exp_x^{-1} \bar{w}, \nabla \phi^i(x) \rangle d\bar{\mu}(x) \right) \quad (\text{from Lemma 2.4}) \\ &= -2 \int_M \langle \exp_x^{-1} \bar{w}, \sum_{i=1}^N \lambda_i \nabla \phi^i(x) \rangle d\bar{\mu}(x) \\ &= 0 \quad (\text{from (3.5)}) \end{aligned} \tag{3.6}$$

Note that $\Phi(t)$ is convex, since from Theorem 2.7, $W_2^2(\delta_{\gamma^i(t)}, \mu^i(t))$ is convex in t . Combined with (3.6), convexity of Φ implies

$$\Phi(0) \leq \Phi(t).$$

But, notice that

$$\begin{aligned} \Phi(0) &= \sum_{i=1}^N \lambda_i W_2^2(\bar{w}, \bar{\mu}) = \text{var}(\bar{\mu}), \\ \Phi(1) &= \sum_{i=1}^N \lambda_i W_2^2(\omega^i, \mu^i) = \int_{P(M)} \text{var}(\mu) d\Omega(\mu). \end{aligned}$$

This establishes the result in the first case.

Case 2: Next, we consider the case when Ω has a unique barycenter.

Noting that Wasserstein space $P(M)$ over M is a Polish space, we can choose a sequence $\Omega^N = \sum_{i=1}^N \lambda_i \delta_{\mu_i}$ of finitely supported measures on $P(M)$ converging weakly-* to Ω , by [20, Theorem 6.18]. For each N , we can also choose at least one of the μ_i to be absolutely continuous with respect to volume, by weak-* density of absolutely continuous measures on M . By Lemma 3.2 and uniqueness, we know that the barycenters $\bar{\mu}^N$ of Ω^N converge weakly to the barycenter $\bar{\mu}$ of Ω . Now, by the above

$$\text{var}(\bar{\mu}^N) \leq \int_{P(M)} \text{var}(\mu) d\Omega^N(\mu).$$

Now, take the limit as $N \rightarrow \infty$. The right hand side tends to $\int_{P(M)} \text{var}(\mu) d\Omega(\mu)$ by Corollary 3.4. The left hand side tends to $\text{var}(\bar{\mu})$ by Lemma 3.3. This completes the proof in the case when the barycenter is unique.

Case 3: Finally, we consider the general case.

Let $\bar{\mu}$ be a (not necessarily unique) barycenter of Ω . By case 2 and Lemma 3.5, we have

$$\text{var}(\bar{\mu}) \leq \frac{1}{2} \text{var}(\bar{\mu}) + \frac{1}{2} \int_{P(M)} \text{var}(\mu) d\Omega(\mu)$$

which easily implies the desired result. \square

Note that the above theorem does not hold if the curvature assumption is removed, as the following example illustrates:

Example 3.7 (Sphere). *Let M be the 2-dimensional round sphere of circumference 2, i.e. the Riemannian distance from the north to south pole is 1. Then, consider the two measures $\mu_0 = \delta_n$, $\mu_1 = \delta_s$, where n and s denote the north and south pole, respectively. Let $\Omega = \frac{1}{2}(\delta_{\mu_0} + \delta_{\mu_1})$ on $P(M)$ and note that $\int_{P(M)} \text{var}(\mu) d\Omega(\mu) = \frac{1}{2} \text{var}(\delta_n) + \frac{1}{2} \text{var}(\delta_s) = 0$. There are infinitely many W_2 -barycenters; namely, any probability measure supported on the equator is a W_2 -barycenter of Ω . In particular, δ_z is a W_2 -barycenter for any z in the equator, which has vanishing variance. This does not violate the inequality of Theorem 3.6. However, uniform measure $\bar{\mu}$ on the equator is also a barycenter. Then note that $\text{var}(\bar{\mu}) = \int d^2(t, n) d\mu(t)$ for the north pole n , and so $\text{var}(\bar{\mu}) = 1/4 > 0 = \int_{P(M)} \text{var}(\mu) d\Omega(\mu)$.*

We close this section by noting a consequence of Theorem 3.6 which will be relevant in the next section.

Corollary 3.8 (Variance gets reduced at the barycenter of an orbit of isometries). *Let G be a set of isometries on a complete simply connected manifold M of nonnegative curvature. Let Ω be a probability measure on G . Consider a probability measure μ on M and assume that there exists a large geodesic ball that contains the union of the supports of the measures $g_{\#}\mu$ for all $g \in G$. Let $\bar{\mu}_{\Omega}$ be a W_2 barycenter of the measure $(g \mapsto g_{\#}\mu)_{\#}\Omega$ on $P(M)$.*

Then,

$$\text{var}(\bar{\mu}_{\Omega}) \leq \text{var}(\mu).$$

Proof. The corollary immediately follows from Theorem 3.6 since $\text{var}(g_{\#}\mu) = \text{var}(\mu)$ for each isometry g . Note that under the nonnegative curvature and simply connected assumption, each geodesic ball is geodesically convex. \square

Remark 3.9. *Although we do not pursue it here, by assuming some decay conditions on the measure Ω on G , as well as considering the space $P_2(M)$ of measures with finite variance, one may extend the above results to non compact cases, in particular, to include isometry group actions on the whole Euclidean space or the hyperbolic space.*

4 Comparison with linear interpolation

In this section, we obtain, as a corollary to Theorem 3.6, a comparison result for the variance functional between the linear barycenters and the W_2 barycenters.

We first consider the linear interpolation between probability measures, $\mu_t = (1-t)\mu_0 + t\mu_1$. We then have that

$$t \mapsto \text{var}(\mu_t) = \min_y \int_M d^2(x, y) d\mu_t(x)$$

is an infimum of affine functions, and hence concave. Define $BC^L(\Omega) \in P(M)$ to be the linear barycenter of the measure Ω on $P(M)$; that is, for each Borel $A \subseteq M$,

$$BC^L(\Omega)[A] := \int_{P(M)} \mu(A) d\Omega(\mu).$$

Then, the (linear) concavity of the variance and the classical (linear) Jensen's inequality implies

$$\text{var}(BC^L(\Omega)) \geq \int_{P(M)} \text{var}(\mu) d\Omega(\mu). \quad (4.1)$$

Note that this holds for any Riemannian manifold M ; the sectional curvature does not play a role. On the other hand, we have:

Corollary 4.1. *Let M be a compact geodesically convex domain in a complete manifold of nonpositive sectional curvature, Ω be a probability measure on $P(M)$, and $\bar{\mu}$ be its W_2 -barycenter. Then,*

$$\text{var}(\bar{\mu}) \leq \text{var}(BC^L(\Omega)).$$

Proof. This immediately follows from the preceding inequality combined with Theorem 3.6. \square

This inequality seems quite intuitive to us. Consider the following, naive explanation. For simplicity, focus on the interpolation between two measures; displacement interpolation (the two measure case of W_2 barycenters) moves the support of one measure to the other continuously along geodesics, so that the support of the interpolant should not be much more spread out than the supports of the two original measures. On the other hand, the support of the linear interpolant is the union of the supports of the two original measures, and so we expect it to be more spread out (i.e., have higher variance) than the displacement interpolant. However, this intuition is somewhat misleading, as it does not require any assumptions on the curvature; as the following example demonstrates, the nonpositive curvature condition in Corollary 4.1 is essential.

Example 4.2 (Balloon on a string). *Consider the sphere S^2 of circumference 1 (the “balloon”), so the distance between the north and south poles is $\frac{1}{2}$, with*

a line segment of length 1 (the “string”) attached to the south pole. Let x be the north pole, and set $x_0 = \exp_x v$, $x_1 = \exp_x(-v)$ for some $v \in T_x M$ with $|v| = \epsilon < \frac{1}{4}$ (that is, x_0 and x_1 are found at the same distance from the north pole, along opposite directions.) Let y be the point on the line segment at a distance $\frac{1}{2} - \epsilon$ from the south pole.

Now, set $\mu_0 = \frac{1}{2}[\delta_y + \delta_{x_0}]$ and $\mu_1 = \frac{1}{2}[\delta_y + \delta_{x_1}]$. Note that the south pole is the W_2 -barycenter of both of these measures, and they each have variance $(\frac{1}{2} - \epsilon)^2$. It is then easy to see that the south pole is also the barycenter of the linear interpolation: $\mu_{1/2}^L := \frac{1}{2}[\mu_0 + \mu_1]$ and that $\text{var}(\mu_{1/2}^L) = (\frac{1}{2} - \epsilon)^2$ as well. On the other hand, the displacement interpolant is given by $\mu_{1/2}^W := \frac{1}{2}[\delta_y + \delta_x]$ (recalling that x is the north pole) whose variance is given by

$$\text{var}(\mu_{1/2}^W) = \frac{d^2(x, y)}{4} = \frac{(1 - \epsilon)^2}{4} > \text{var}(\mu_{1/2}^L).$$

Although the metric space in the example is not a smooth manifold, it can easily be smoothed out to construct smooth examples where the preceding variance inequality holds.

5 W_2 Projection to the G -invariance set

In this section, we consider isometry group actions on the underlying space M , which also induce isometry group actions on $P(M)$. We are interested in the Wasserstein W_2 -barycenter of the orbit of the group action, in relation to functionals on $P(M)$. Our focus in this section is on the variance functional for nonpositively curved underlying space, so that we can use the results of the preceding sections. But, the similar results hold for other examples; see Remark 5.4 and Example 5.5.

We begin by showing that the projection onto the invariance set coincides with the barycenter of the orbit under left Haar measure.

Proposition 5.1 (Projection to G -invariance set). *Let G be a group of isometries on a Riemannian manifold M and H be a left invariant probability measure on G (here, the group G has to be compact). For a given probability measure $\mu \in P(M)$, assume that the barycenter $BC_G^W(\mu)$ of $\Omega_\mu = (\mu \mapsto g_\# \mu)_\# H$ is unique. Define the G -invariant set $I_G = \{\nu \in P(M) \mid g_\# \nu = \nu, \forall g \in G\}$. Then,*

1. $BC_G^W(\mu) \in I_G$.
2. $BC_G^W(\mu)$ is the unique W_2 projection of μ to I_G ; that is $\{BC_G^W(\mu)\} = \arg\min_{\nu \in I_G} W_2^2(\nu, \mu)$, or, using the notation in the introduction,

$$BC_G^W(\mu) = P_G^W(\mu).$$

This should be a well-known standard fact from metric geometry, but, we give its proof for completeness.

Remark 5.2. *The uniqueness condition on the barycenter is satisfied when μ is absolutely continuous with respect to volume (as then each $g_{\#}\mu$ is clearly absolutely continuous as well) by Proposition 3.1.*

Proof. For simplicity of notation, we denote $\bar{\mu} = BC_G^W(\mu)$. We prove the two assertions below:

1. For each $g, g' \in G$, by the isometry property, we have $W_2(g_{\#}\bar{\mu}, gg'_{\#}\mu) = W_2(\bar{\mu}, g'_{\#}\mu)$. Thus,

$$\begin{aligned} \int_G W_2^2(\bar{\mu}, g'_{\#}\mu) d\Omega(g') &= \int_G W_2^2(g_{\#}\bar{\mu}, gg'_{\#}\mu) d\Omega(g') \\ &= \int_G W_2^2(g_{\#}\bar{\mu}, gg'_{\#}\mu) d\Omega(gg') \quad (\text{as } \Omega \text{ is left invariant}) \end{aligned}$$

This implies that $g_{\#}\bar{\mu}$ is a barycenter of Ω ; by the uniqueness assumption, this shows $\bar{\mu} = g_{\#}\bar{\mu}$. As this holds for each $g \in G$, we have $\bar{\mu} \in I_G$.

2. Notice that if $\nu \in I_G$, then, $W_2(\nu, g_{\#}\mu) = W_2(\nu, \mu)$ for all $g \in G$. Thus,

$$\begin{aligned} W_2^2(\nu, \mu) &= \int_G W_2^2(\nu, \mu) d\Omega(g) \\ &= \int_G W_2^2(\nu, g_{\#}\mu) d\Omega(g) \\ &\geq \int_G W_2^2(\bar{\mu}, g_{\#}\mu) d\Omega(g) \quad (\text{from the definition of } \bar{\mu}) \\ &= \int_G W_2^2(\bar{\mu}, \mu) d\Omega(g) \quad (\text{since } \bar{\mu} \in I_G \text{ from 1.}) \\ &= W_2^2(\bar{\mu}, \mu) \end{aligned}$$

This shows that $\bar{\mu}$ is the minimum of $\{W_2^2(\nu, \mu)\}_{\nu \in I_G}$. Noting that the inequality is strict unless $\nu = \bar{\mu}$, by the uniqueness of the barycenter, completes the proof. \square

An interesting consequence follows:

Corollary 5.3 (Variance gets reduced at the W_2 projection to the invariant set). *Under the same notation as in Proposition 5.1, assume further that M is a complete, simply connected nonpositively curved manifold. For each absolutely continuous probability measure μ with compact support,*

$$\text{var}(P_G^W \mu) \leq \text{var}(\mu).$$

Proof. This immediately follows from Corollary 3.8 and Proposition 5.1. \square

If μ is absolutely continuous, with a density f in L^2 , it is straightforward to see that the linear barycenter $BC_G^L(\mu)$ of Ω_μ is absolutely continuous as well, and that its density is given by

$$\bar{f}(x) = \int_G f(g^{-1}(x)) dH(g).$$

This \bar{f} is the L^2 minimizer of the functional

$$h \mapsto \int_G \|h - f \circ g^{-1}\|^2 dH(g)$$

and it also coincides with the L^2 projection $P_G^{L^2}(\mu)$ of μ onto the subspace of L^2 functions which are invariant under the action of G . From (4.1) and Theorem 5.3, we have

$$\text{var}(P_G^W(\mu)) \leq \text{var}(\mu) \leq \text{var}(P_G^{L^2}(\mu)). \quad (5.1)$$

In particular, if we consider the case $G = SO(n)$ acting on the Euclidean space \mathbb{R}^n (or any rotationally symmetric nonpositively curved metric on \mathbb{R}^n , e.g. the hyperbolic metric), projecting onto the invariant set can be interpreted as finding the best rotationally invariant approximation of μ . Finding the best approximation of a measure μ , in the Wasserstein sense, by a radially symmetric measure decreases the variance. On the other hand, finding the best approximation of a measure by a radially symmetric measure in the L^2 sense *increases* the variance.

Remark 5.4. *It is worth noting that Corollary 5.3 holds whenever the variance is replaced with any functional F which is convex over W_2 -barycenters (including the three main types discovered in [11], whose convexity over W_2 barycenters are obtained in [1] on the Euclidean space and in [8] on Riemannian manifolds with nonnegative Ricci curvature, and more generally on smooth metric measure spaces satisfying the $CD(K, N)$ condition for $K \geq 0$), provided the functionals are invariant under an isometry group G ; that is, provided $F(A\#\mu) = F(\mu)$ for all μ and all $A \in G$. We feel the variance case on Hadamard manifolds (another name for complete, simply connected nonpositively curved manifolds) is of special interest as the opposite holds true for linear barycenters. Linear and Wasserstein projections give two ways to canonically generate a G -invariant measure from a given measure; one of these decreases the variance while the other increases it.*

Example 5.5. *As an illustrative example, consider the entropy functional, $F(f d\text{vol}) = \int_M f(x) \ln(f(x)) d\text{vol}(x)$ on a manifold with nonnegative Ricci curvature. It is well known that F is displacement convex; our recent work [8] extends this result to show that F is in fact convex over barycenters. For any absolutely continuous measure μ and any compact group of isometries G on M , we then get*

$$F(BC_G^W(\mu)) \leq F(\mu). \quad (5.2)$$

In the particular case when $M = S^n$ is the round sphere and G is the whole rotation group, the barycenter $BC_G^W(\mu) = \text{vol}/\text{vol}(S^n)$ must be uniform measure, as this is the only probability measure on S^n which is invariant under this group. It is well known that uniform measure minimizes the entropy, so this is consistent with (5.2). For smaller rotation groups G , symmetrizing with respect to G (that is, projecting onto the G -invariant set) reduces the entropy, by (5.2).

References

- [1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. SIAM J. Math. Anal., 43(2):904–924, 2011.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics, ETH Zürich. Birkhäuser Verlag, Basel, 2005.
- [3] Y. Brenier. Decomposition polaire et rearrangement monotone des champs de vecteurs. C.R. Acad. Sci. Paris. Ser. I Math., 305:805–808, 1987.
- [4] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality a la Borell, Brascamp and Lieb. Invent. Math., 146(2):219–257, 2001.
- [5] Alfred Galichon and Nassif Ghoussoub. Variational representations for N -cyclically monotone vector fields. Pacific J. Math., 269(2):323–340, 2014.
- [6] Nassif Ghoussoub and Bernard Maurey. Remarks on multi-marginal symmetric Monge-Kantorovich problems. Discrete Contin. Dyn. Syst., 34(4):1465–1480, 2014.
- [7] Nassif Ghoussoub and Abbas Moameni. Symmetric Monge-Kantorovich problems and polar decompositions of vector fields. Geom. Funct. Anal., 24(4):1129–1166, 2014.
- [8] Y.-H. Kim and B. Pass. Wasserstein barycenters over Riemannian manifolds . Preprint. Currently available at arXiv:1412.7726.
- [9] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. Ann. of Math. (2), 169(3):903–991, 2009.
- [10] R. McCann. Polar factorization of maps on Riemannian manifolds. Geom. Funct. Anal., 11:589–608, 2001.
- [11] R.J. McCann. A convexity theory for interacting gases and equilibrium crystals. Adv. Math., 128(1):153–179, 1997.
- [12] Abbas Moameni. Invariance properties of the Monge-Kantorovich mass transport problem. Preprint at arXiv:1311.7051.
- [13] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal., 173(2):361–400, 2000.
- [14] Brendan Pass. Optimal transportation with infinitely many marginals. J. Funct. Anal., 264(4):947–963, 2013.
- [15] Brendan Pass. Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions. Discrete Contin. Dyn. Syst., 34(4):1623–1639, 2014.

- [16] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. Acta Math., 196(1):65–131, 2006.
- [17] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. Acta Math., 196(1):133–177, 2006.
- [18] K.T. Sturm. Probability measures on metric spaces of nonpositive curvature. Contemp. Math., 338:357–390, 2003.
- [19] C. Villani. Topics in optimal transportation, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.
- [20] C. Villani. Optimal transport: old and new, volume 338 of Grundlehren der mathematischen Wissenschaften. Springer, New York, 2009.
- [21] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. Comm. Pure Appl. Math., 58(7):923–940, 2005.
- [22] Danila Zaev. On the Monge-Kantorovich problem with additional linear constraints. Preprint at arXiv:1404.4962.